BUAN 3210 Quick Project 1
Professor Green Due 5:59 PM Sat, April 25 (up to 12hrs late for full credit)

You are interviewing for a job at Zillow. Jennifer Lee, the recruiter, has provided you a data set on 2014-15 home sales in King County that she would like you to analyze, find several insights on the market, and write up your findings. This is effectively a test of your basic analysis and communication skills. Jennifer says you can work on this with one of the other job applicants because they want to see your ability to work in teams. She understands that you do not know much about the housing market so may need to bounce ideas off of each other. The specifics of the task are outlined below. Jennifer will be sharing your findings with the Zillow hiring manager, so it is critical that your writing and graphs in the memo are clear to people without a technical background and concise for someone that has limited time.

The data set "kc_home_sales.csv" contains the following variables:

1. date: date of the sale
2. price: sale price of the home
3. bedrooms: number of bedrooms in the home
4. bathrooms: number of bathrooms in the home
5. sqft_living: house size in square feet, not including garage
6. sqft_lot: lot size in total square feet
7. floors: numer of levels in the house, x.5 means there is a split-level
8. waterfront: lot is located on the water
9. view: home has a view, higher number means better view
10. grade: quality of initial home build and finishing's, higher is better
11. condition: the overall maintenance of the home, regardless of initial build quality or grade
12. yr_built: year the home was built
13. yr_renovated: most recent year the home was renovated, 0 if not renovated
14. zip_code: the zip code of the home's neighborhood
15. lat: latitude location of the home
16. long: longitude location of the home

Each observation is for the sale of a specific home

Your Task: You may work alone or in groups of two (no larger groups)

1. Complete a base and detailed EDA in an R Notebook as your Technical Appendix. The basic EDA is what we have discussed in class, univariate/multivariate/non-graphical/graphical EDA. A detailed EDA is when you dig deeper into the data to answer questions you raised in the base EDA. Use a ***variety*** of tidyverse data verbs to select, filter, group, mutate, and summarize variables—be adventurous! Create exploratory and presentation visualizations. Knit the Technical Appendix with a table of contents to show all code and output. Make sure your code is well documented and include data comments/questions regarding your findings. Remember code documentation is in the code chunk and data comments are after the output (but leave a blank line between output and data comment so they begin on a new line). You may want to label sections by EDA step and/or the question you are exploring to organize the TA. Save visuals so you can load them in your memo, as demonstrated in week 2 of class.

2. Determine your two preferred market insights for the King County housing market and create presentation quality visuals to support each finding. Though you may create many graphs in your EDA, you will only want to use a several of your visuals to illustrate your

findings. It is important the visuals are easy to understand so make sure it is not complicated or cluttered, with clear titles and axis labels. I expect you to have at least two nicely formatted visuals; quality is more important than quantity. Write up a two-page, four-paragraph summary of the project:

a. An introduction: why is the analysis important, a brief description of the data, and a brief description of your findings. Think of this as an executive summary.
b. A paragraph and presentation quality visual(s) **for each of the two findings**. If you have more than two findings then you have to decide which you think are the most important. You can mention other findings in a sentence or two in the conclusion/summary.
c. A conclusion/summary.

Knit these as an html or Word memo with text and visuals (no code) from RMarkdown (use my examples from week 2 as a template). Write the memo for a lay audience—no code, no technical visuals (boxplots and correlation graphs are technical so do not use them), and no jargon. The memo with visuals should be 2 pages or less (html does not have page breaks, so you will have to use your judgement).

Submit the technical appendix and memo in html or Word to Canvas (not Rmd files).

Quick Project Grade Rubric

Category - Percent of grade	Score	**Comments**
Introductory Paragraph – (3) Compelling (1), data (1), brief finding (1)		
Finding 1 – (4) Visual (2.5), clear description (1.5)		
Finding 2 – (4) Visual (2.5), clear description (1.5)		
Conclusion – (2) Summary (1), complete (1)		
Technical Appendix – (12) Clear organization (1), documentation of code (2), data questions (2), basic EDA (5), detailed EDA (2)		
Total	**/25**	

What does “clear description” mean? When you are writing to a lay audience you need to assume they have only a limited background in quantitative methods. You can generally assume that people know what an average is and understand that variance or standard deviation relates to variation of the data, but not much more. They may or may not know what the median is and why it might be better than the mean at representing data. You would likely need to explain what a box plot or histogram illustrates. When trying to determine if terminology or a visualization is “clear” or not, think about saying or showing it to a friend without a quantitative background. Would they know what the terminology means, or would you need to provide more background for them to understand it? In fact, a good way to determine if your memo is written appropriately is to have a non-technical person read your memo. Clear description also means your writing is professional – no slang, typos, and good grammar. The Seattle University writing center is a good group to work with.